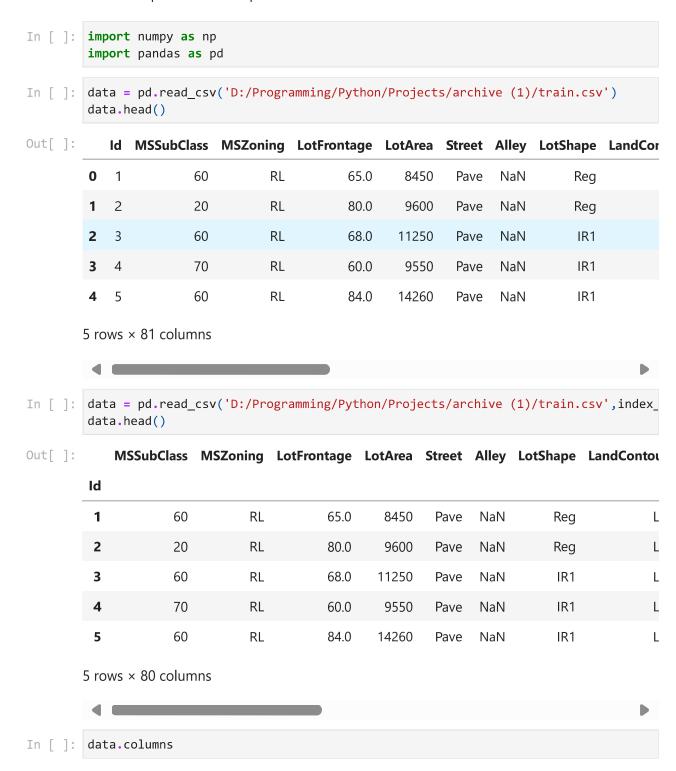
### **House Price Prediction Project**

## 1. Problem definition

Goal: predict the sale price for each house



#### 2. Feature selections

- Choose features to train ML model
- Need to use "Feature Engineering" to identify features needed

```
In [ ]: features = ['LotArea', 'YearBuilt', '1stFlrSF', '2ndFlrSF', 'FullBath', 'BedroomAbvGr
```

## 3. Spliting datasets

• "data": dataset

Ou+ [ ] .

- "X": data[features]
- "y": target variable "SalePrice"

```
In [ ]: X = data[features]
y = data['SalePrice']
data.head()
```

out[ ]:		MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandConto
	Id								
	1	60	RL	65.0	8450	Pave	NaN	Reg	L
	2	20	RL	80.0	9600	Pave	NaN	Reg	L
	3	60	RL	68.0	11250	Pave	NaN	IR1	L
	4	70	RL	60.0	9550	Pave	NaN	IR1	L
	5	60	RL	84.0	14260	Pave	NaN	IR1	L

5 rows × 80 columns

```
In []: X.head()
```

Out[]:		LotArea	YearBuilt	1stFlrSF	2ndFlrSF	FullBath	BedroomAbvGr	TotRmsAbvGrd		
	Id									
	1	8450	2003	856	854	2	3	8		
	2	9600	1976	1262	0	2	3	6		
	3	11250	2001	920	866	2	3	6		
	4	9550	1915	961	756	1	3	7		
	5	14260	2000	1145	1053	2	4	9		
In [ ]:	y.head()									
Out[ ]:	Id 1 2 3 4 5 Nar	208500 181500 223500 140000 250000 me: SaleP		pe: int64						
In [ ]:	fro	<pre>X, y&gt; X_train, y_train, X_test, y_test  from sklearn.model_selection import train_test_split</pre>								
In [ ]:		<pre>X_train, X_test, y_train, y_test = train_test_split(X,y, train_size=0.8, test_s X_train.shape</pre>								
Out[ ]:	(1:	(1168, 7)								
In [ ]:	X_t	X_test.shape								
Out[ ]:	(292, 7)									

# 4. Training machine learning model

Predict